

Translation of Abstract for German Patent DE 43 39 161 C2

**(54) Method for Tracking the Movements of a Human Hand in a Chronological Sequence of Digital Color Images**

(57) A method for tracking movements of a human hand in a chronological sequence of digital color images, for which a segmenting of the hand is carried out with the aid of a color histogram of this hand and for which a group of image points is classified as belonging to the hand segment if a comparison of the color histogram for this group of image points with the color histogram of the hand sufficiently coincides.

**This Page Blank (uspto)**

19 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

12 Patentschrift  
10 DE 43 39 161 C 2

51 Int. Cl. 6:  
G 06 T 7/20  
G 06 K 9/62

21 Aktenzeichen: P 43 39 161.3-53  
22 Anmeldetag: 18. 11. 93  
43 Offenlegungstag: 18. 2. 95  
45 Veröffentlichungstag  
der Patenterteilung: 28. 11. 98



DE 43 39 161 C 2

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

30 Unionspriorität: 32 33 31  
09.08.93 EP 93112720.3

73 Patentinhaber:  
Siemens AG, 80333 München, DE

72 Erfinder:  
Ahmad, Subutal, Dr., 81675 München, DE

56 Für die Beurteilung der Patentfähigkeit  
in Betracht gezogene Druckschriften:

DE 43 09 878 A1  
US 51 68 531  
US 50 18 173

KRUEGER, Myron W.: »Artificial Reality II«,  
Addison-Verlag Publishing Company, 1991,  
S. 170-182;  
6th International Conference on Pattern Recognition,  
München 1982, S. 520-522;

SEGEN, J.: »Model Learning and Recognition of  
Nonrigid Objects« IEEE Computer Society Confer-  
ence on Computer Vision and Pattern Recognition, San  
Diego, CA, 1989;  
WIRTZ/MAGGIONI: »Image Glove: A Novel Way to  
Control Virtual Environments« Proceedings of Virtual  
Reality Systems, 1993, New York;  
SWAIN/BALLARD: »Color Indexing« Int. Journal of  
Computer Vision, 7:1, 11-32, 1991;

54 Verfahren zur Verfolgung von Bewegungen einer menschlichen Hand in einer zeitlichen Folge digitaler  
Farbbilder

57 Verfahren zur Verfolgung von Bewegungen einer mensch-  
lichen Hand in einer zeitlichen Folge digitaler Farbbilder, bei  
dem eine Segmentierung der Hand mit Hilfe eines Farbhisto-  
gramms dieser Hand durchgeführt wird, und bei dem eine  
Gruppe von Bildpunkten als zum Handsegment gehörend  
klassifiziert wird, falls ein Vergleich des Farbhistogramms  
dieser Gruppe von Bildpunkten mit dem Farbhistogramm der  
Hand eine hinreichende Übereinstimmung ergibt.

DE 43 39 161 C 2

## Beschreibung

Die Erfindung betrifft ein Verfahren zur Verfolgung von Bewegungen einer menschlichen Hand in einer zeitlichen Folge digitaler Farbbilder. Derartige Verfahren werden z. B. zur Steuerung von Datenverarbeitungssystemen oder daran angeschlossenen Systemen, z. B. in der Medizintechnik, der Fertigungstechnik oder der Automatisierungstechnik benötigt. In der Literatur sind verschiedene Verfahren dieser Art beschrieben worden. So beschreibt Krueger in [1] ein System, welches die Umrisse bzw. die Umrandungskurve einer Hand in zwei Dimensionen detektieren kann, unter der Voraussetzung eines gleichförmigen Hintergrundes hinter der Hand. Segen beschreibt in [2] ein System mit ähnlichen Eigenschaften, welches die Bewegungen einer Hand in zwei Dimensionen unter der Annahme eines weißen Hintergrundes verfolgen kann; dieses System ist auch bereits dazu in der Lage, eine kleine Anzahl von Gesten zu erlernen. Schließlich beschreiben Wirtz und Maggioni in [3] ein System zur Verfolgung von Handbewegungen in drei Dimensionen, das zwar unabhängig von den Eigenschaften des Bildhintergrundes arbeitet, dafür aber voraussetzt, daß die betreffende Person einen Handschuh mit darauf angebrachten speziellen Markierungen trägt.

Keines der bekannten Verfahren ist jedoch in der Lage, unabhängig von den Eigenschaften des Bildhintergrundes und in drei Dimensionen zu arbeiten, ohne auf spezielle Markierungen auf der Handoberfläche angewiesen zu sein. Aus der amerikanischen Patentschrift US 5 168 531 ist ein Verfahren zur Verfolgung von Bewegungen einer menschlichen Hand einer zeitlichen Folge digitaler Bilder, die gegebenenfalls auch Farbbilder sein können, bekannt. Dabei werden aufgenommene Bilddaten auf das Vorhandensein elementarer Merkmale hin analysiert und mit vordefinierten Merkmalen verglichen. Ferner ist aus der amerikanischen Patentschrift US 5 016 173 bekannt, die Segmentierung von Farbbildern der menschlichen Haut mittels Farbhistogrammen durchzuführen. Dieses Verfahren ist jedoch kompliziert und damit nur im Vergleich zu dem erfindungsgemäßen Verfahren langsam durchführbar.

Der Erfindung liegt das Problem zugrunde, ein Verfahren zur Verfolgung von Bewegungen einer menschlichen Hand in einer zeitlichen Folge digitaler Farbbilder anzugeben, welches in der Lage ist, dreidimensionale Informationen auszuwerten, unabhängig von den Eigenschaften des Bildhintergrundes arbeitet, nicht auf spezielle Markierungen auf der Handoberfläche angewiesen ist, und einfacher und somit schneller durchführbar ist als die bekannten Verfahren.

Dieses Problem wird erfindungsgemäß durch ein Verfahren zur Verfolgung von Bewegungen einer menschlichen Hand in einer zeitlichen Folge digitaler Farbbilder mit Merkmalen nach Anspruch 1 gelöst.

Bei diesem Verfahren wird eine Segmentierung der Hand mit Hilfe eines Farbhistogramms dieser Hand durchgeführt.

Vorteilhafte Weiterbildungen der Erfindung ergeben sich aus den Unteransprüchen.

Eine bevorzugte Ausführungsform des erfindungsgemäßen Verfahrens klassifiziert dazu eine Gruppe von Bildpunkten als zum Handsegment gehörend, falls ein Vergleich des Farbhistogramms dieser Gruppe von Bildpunkten mit dem Farbhistogramm der Hand eine hinreichende Übereinstimmung ergibt.

Diese Vorgehensweise ist mit dem Vorteil einer geringeren Rauschempfindlichkeit der Segmentierung verbunden. Bei einer weiteren bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens werden aus den Koordinaten der Mittelpunkte der zum Handsegment gehörenden Gruppen von Bildpunkten die Koordinaten des Schwerpunktes des Handsegments, die Momente zweiter Ordnung um diesen Schwerpunkt und daraus die Orientierung der Hauptachse des Handsegments ermittelt, und es wird die senkrechte zur Bildebene gemessene momentane Entfernung der Hand von der Kamera durch einen Vergleich der momentanen Fläche des Handsegments mit seiner Fläche im Zeitpunkt einer Kalibrierung ermittelt. Da das Moment der Ordnung 0 gerade die Fläche des Handsegments angibt, bietet diese Vorgehensweise einen vorteilhaften Weg zur Bestimmung der Fläche des Handsegments.

Bei einer weiteren bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens werden zur Rauschunterdrückung die ermittelten Meßwerte für die Schwerpunktskoordinaten und die Orientierung der Hauptachse des Handsegments zur zeitlich rekursive Filterung geglättet. Auf diese Weise wird das erfindungsgemäße Verfahren weniger anfällig gegen Rauschen in den gemessenen Schwerpunktskoordinaten.

Bei einer weiteren bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens wird die Segmentierung eines Bildes auf einen Ausschnitt der Bildebene beschränkt, welcher das Handsegment des vorhergehenden Bildes zusammen mit einer geeignet gewählten Umgebung des Handsegments enthält. Auf diese Weise läßt sich eine beachtliche Beschleunigung des erfindungsgemäßen Verfahrens erzielen, welche einer Ausführbarkeit dieses Verfahrens in Echtzeit zugute kommt.

Bei einer weiteren bevorzugten Ausführungsform des erfindungsgemäßen Verfahrens wird die Segmentierung mit Hilfe einer Unterabtastung des zu segmentierenden Bildes durchgeführt. Auch diese Vorgehensweise ist mit einer bedeutenden Beschleunigung des erfindungsgemäßen Verfahrens verbunden.

Im folgenden wird die Erfindung anhand bevorzugter Ausführungsbeispiele näher beschrieben.

Zur Durchführung des erfindungsgemäßen Verfahrens wird eine menschliche Hand und ihre Bewegungen mit Hilfe einer Kamera, welche eine zeitliche Folge digitaler Farbbilder erzeugt, aufgenommen. Liegt diese zeitliche Folge von digitalen Farbbildern vor, besteht eine erste Aufgabe eines Systems zur Durchführung des erfindungsgemäßen Verfahrens in der Trennung der zu verfolgenden Hand von dem Bildhintergrund. Das hauptsächliche Problem bei der Entwicklung eines entsprechenden Moduls zur Segmentierung besteht darin, daß diese Segmentierung unempfindlich gegenüber normalerweise auftretenden Bildhintergrundformen sein muß und daß das Verfahren in Echtzeit ablaufen muß. Dies ist deshalb keine einfache Aufgabe, weil das Bild im allgemeinen eine Vielzahl irrelevanter und verwirrender Details enthalten kann. Die grundlegende Idee die den Kern des erfindungsgemäßen Verfahrens bildet, besteht darin, die spezielle Farbe einer menschlichen Hand dazu zu verwenden, diese Hand vom Bildhintergrund zu trennen. Die meisten Objekte in üblichen Arbeitsumgebungen haben

nicht die gleiche Farbe wie eine menschliche Hand, so daß eine Segmentierung aufgrund der Farbunterschiede verhältnismäßig saubere Ergebnisse liefern sollte. Bei dem erfindungsgemäßen Verfahren wird ein Farbhistogramm dazu benutzt die Verteilung der Farben in einem Flächenausschnitt aus einem Bild der menschlichen Haut zu schätzen. Sobald diese Schätzung vorliegt, kann sie dazu verwendet werden zu entscheiden, welche Teile eines Bildes zur Handfläche gehören und welche nicht dazu gehören.

Die grundlegende Idee hinter dem Begriff "Farbhistogramm" ist eine Unterteilung des RGB (Rot-Grün-Blau) Farbraumes in einer Anzahl von Zellen. Zur Gewinnung eines repräsentativen Farbhistogramms wird ein Ausschnitt eines Bildes einer Hautoberfläche ausgewählt. Aus der Menge der Bildpunkte dieses Ausschnittes wird ein Farbhistogramm gebildet, in dem jeweils die Zahl derjenigen Bildpunkte ermittelt wird, deren Farbwert in einer gegebenen Zelle des Farbraums liegt. Anstelle des RGB Raums könnte auch ein normalisierter zweidimensionaler Farbraum verwendet werden, welcher dazu beiträgt, die Auswirkungen von Beleuchtungsänderungen abzuschwächen.

Ist ein Beispielbildpunkt mit Hautfarbe gegeben, dessen Farbwerte mit  $r$ ,  $g$  und  $b$ , bezeichnet werden, können dessen normalisierte Farben nach den Formeln

$$r' = r/(r+g+b+1)$$

und

$$b' = b/(r+g+b+1)$$

berechnet werden. Die Werte für  $r'$  und  $b'$  liegen zwischen 0 und 1 und können als relative Anteile der Farben Rot und Blau interpretiert werden. Jede der Achsen  $r'$  und  $b'$  im zweidimensionalen normalisierten Farbraum kann mit einem Faktor  $d$  (z. B.  $d = 5$ ) diskretisiert werden. Das normalisierte Farbhistogramm ist dann eine  $d \times d$ -Matrix. Für jeden Bildpunkt kann der Index jeder Dimension dann nach den Beziehungen

$$i = [r'd] \text{ und } j = [b'd]$$

berechnet werden.

Um das Histogramm zu erstellen wird die Matrix zunächst mit Nullen initialisiert. Dann wird für jeden Bildpunkt  $(x, y)$  in dem Satz von Bildpunkten der zum Training des Histogramms verwendet wird die normalisierten Farbwerte berechnet welche dem Histogrammindex entsprechen, und der entsprechende Histogrammwert an dieser Stelle wird um 1 erhöht:  $H(i, j) \leftarrow H(i, j) + 1$ . Durch diese Art der Histogrammberechnung von Hautproben wird im wesentlichen eine grobe Schätzung der Wahrscheinlichkeitsverteilung der Farben in der Hautoberfläche ermittelt.

Zur Laufzeit vergleicht das erfindungsgemäße Verfahren laufend kleine Ausschnitte des momentanen digitalen Bildes (Gruppen von Bildpunkten) mit dem gespeicherten Histogramm, wobei eine der üblichen Algorithmen zum Vergleich von Histogrammen verwendet werden kann. Ein verbreiteter Weg zum Vergleich zweier Histogramme ist der Algorithmus von Swain und Ballard [4]. Bei diesem Algorithmus wird die Vergleichszahl  $M$  zwischen den Histogrammen  $p$  und  $q$  definiert als:

$$(1) \quad M_{p,q} = \frac{\sum_{i,j} \min(H^p(i, j), H^q(i, j))}{\sum_{i,j} H^p(i, j)}$$

Diejenigen Gruppen von Bildpunkten (Bildausschnitte) deren Vergleichszahl über einem vorgegebenen Schwellwert (typisch 0,9) liegt werden als Bereiche mit Hautfarbe angesehen und es wird davon ausgegangen, daß diese Bereiche zur Hand gehören.

Sobald das gesamte Bild auf diese Weise segmentiert wurde, liegt eine Liste von Bildausschnitten des Bildes vor, welche zu dem gespeicherten Histogramm passen. Jeder Bildausschnitt besitzt ein Zentrum und eine Fläche. Aus der Menge dieser Bildausschnitte und aus ihren Parametern (Schwerpunkten und Flächen) ist es möglich den Schwerpunkt und die Orientierung der Handfläche zu ermitteln. Dies wird weiter unten beschrieben. Wenn sich die Hand bewegt, ist es auf diese Weise möglich ihre Position und Orientierung im Raum in Echtzeit zu verfolgen.

Nach der Segmentierung kann die Position der Hand im Raum und ihre Orientierung (Anordnung) ermittelt werden. Die Schwerpunkte der segmentierten Gruppen von Bildpunkten (der segmentierten Bildausschnitte) ist gleich der zweidimensionalen Position der Hand:

$$C_x = \frac{1}{N} \sum_i p_{ix}$$

und

$$C_y = \frac{1}{N} \sum_i p_{iy}$$

wobei N die Gesamtzahl der Bildausschnitte bezeichnet die bei der Segmentierung als zur Hand gehörig ermittelt wurden, und wobei  $p_{ix}$  und  $p_{iy}$  die x- und y-Koordinaten des Schwerpunktes des Bildausschnittes mit dem Index i bezeichnen.

Die ebene Drehung (Drehung um die Kameraachse) der Hand wird durch Anpassung einer Ellipse an die segmentierten Bildausschnitte ermittelt, wobei der Winkel und die Hauptachsen der Ellipse ermittelt werden. Dies kann mit Hilfe der Momente geschehen, insbesondere der Momente zweiter Ordnung der Bildausschnitte:

$$(2) \quad m_{2,0} = \sum_i (C_x - p_{ix})^2$$

$$(3) \quad m_{0,2} = \sum_i (C_y - p_{iy})^2$$

$$(4) \quad m_{1,1} = \sum_i (C_y - p_{iy})(C_x - p_{ix})$$

Die Orientierung der Hauptachse ist durch

$$(5) \quad \theta = \frac{1}{2} \tan^{-1} \left( \frac{m_{2,0} - m_{0,2}}{2m_{1,1}} \right)$$

gegeben.

Da das Verfahren mit einer Kamera arbeitet, ist es nicht ganz leicht eine genaue Schätzung des Abstandes der Hand von der Kamera zu erhalten. In jedem Falle ist es jedoch möglich, die relative Tiefe (d. h. die Tiefe relativ zu einer gegebenen Standardanordnung) zu berechnen. Mit dieser Information ist es dann möglich zu sagen, ob die Hand sich der Kamera nähert oder ob sie sich von der Kamera wegbewegt. Eine Methode dies zu ermitteln besteht darin, einfach die Gesamtfläche der segmentierten Bildausschnitte zu ermitteln. Wenn die zu verfolgende Hand der Kamera näher kommt, wird sie im allgemeinen einen größeren Ausschnitt in der Bildebene belegen und die Gesamtfläche aller zur Hand gehörigen Bildausschnitte wird zunehmen. Dieser Effekt kann aber auch durch Störungen, welche durch vorübergehend auftretende Rauschsegmente veranlaßt sind auftreten. Eine robustere Schätzung erhält man deshalb durch eine gewichtete Berechnung der Fläche der Bildausschnitte mit Hilfe einer Gaußkurve, die um den Schwerpunkt der Bildausschnitte zentriert ist:

$$(6) \quad A = A_p \cdot \sum_i \exp \left( \frac{-(C_y - p_{iy})^2 - (C_x - p_{ix})^2}{S} \right)$$

Hierbei bezeichnet  $A_p$  die Fläche die von jedem Bildausschnitt belegt wird und S ist ein Skalierungsfaktor. Um das System zu kalibrieren, muß die Fläche des ersten Bildes gespeichert werden (Kalibrierungsschritt). Anhand dieser Fläche ist es dann bei nachfolgenden Bildern möglich, eine Schätzung des Abstandes der Hand von der Kamera durch Vergleich der momentanen Fläche mit der kalibrierten Fläche zu erhalten. Die z-Koordinate der Schwerpunktslage ergibt sich somit als Quotient der anfänglichen Fläche zur momentanen Fläche. Dieser Wert ist dann ein, wenn die zu verfolgende Hand sich gerade an der gleichen Position befindet wie im zur Kalibrierung verwendeten Bild. Der Wert wird kleiner, wenn die Hand sich der Kamera nähert und größer, wenn sie sich fortbewegt. Auf diese Weise gibt dieser Wert eine Schätzung der relativen Lage der zu verfolgenden Hand in

Bildpunkten bei der Segmentierung belegt, der Unterabtastung angepaßt. Hierdurch bleibt die Berechnung der Tiefe von der Änderung der Unterabtastrate unbeeinflußt.

Im Rahmen dieser Patentanmeldung wurden die folgenden Veröffentlichungen zitiert:

- [1] Krüger, M. (1991), Artificial Reality II. Addison-Wesley.
- [2] Segen, J. (1989), "Model learning and recognition of nonrigid objects" in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA.
- [3] Wirtz, B. and C. Maggioni (1993), Imageglove: "A novel way to control virtual environments", in Proceedings of Virtual Reality Systems '93, New York.
- [4] Swain, M. and D. Ballard (1991), Color indexing "International Journal of Computer Vision 7", 11 — 32.

#### Patentansprüche

1. Verfahren zur Verfolgung von Bewegungen einer menschlichen Hand in einer zeitlichen Folge digitaler Farbbilder, bei dem eine Segmentierung der Hand mit Hilfe eines Farbhistogramms dieser Hand durchgeführt wird, und bei dem eine Gruppe von Bildpunkten als zum Handsegment gehörend klassifiziert wird, falls ein Vergleich des Farbhistogramms dieser Gruppe von Bildpunkten mit dem Farbhistogramm der Hand eine hinreichende Übereinstimmung ergibt.

2. Verfahren nach Anspruch 1, bei dem aus den Koordinaten der Mittelpunkte der zum Handsegment gehörenden Gruppen von Bildpunkten die Koordinaten des Schwerpunktes des Handsegments, die Momente zweiter Ordnung um diesen Schwerpunkt und daraus die Orientierung der Hauptachse des Handsegments ermittelt werden, und bei dem die senkrecht zur Bildebene gemessene momentane Entfernung der Hand von der Kamera durch einen Vergleich der momentanen Fläche des Handsegments mit seiner Fläche im Zeitpunkt einer Kalibrierung ermittelt wird.

3. Verfahren nach Anspruch 2, bei dem zur Rauschunterdrückung die ermittelten Meßwerte für die Schwerpunktskoordinaten und die Orientierung der Hauptachse des Handsegments durch zeitlich rekursive Filterung geglättet werden.

4. Verfahren nach einem der vorhergehenden Ansprüche, bei dem die Segmentierung eines Bildes auf einen Ausschnitt der Bildebene beschränkt wird, welcher das Handsegment des vorhergehenden Bildes zusammen mit einer geeignet gewählten Umgebung dieses Handsegments enthält.

5. Verfahren nach einem der vorhergehenden Ansprüche, bei dem die Segmentierung mit Hilfe einer Unterabtastung des zu segmentierenden Bildes durchgeführt wird.

Abstand zur Kamera an.

Um in diesen Schätzungen ein Rauschen zu unterdrücken ist es vorteilhaft, eine gleitende Mittelwertbildung in Form einer zeitlich rekursiven Filterung der oben angegebenen Parameter zu verwenden. Hierbei werden frühere Schätzungen mit laufenden Messungen so kombiniert, wie dies in den nachstehenden Formeln angegeben ist:

$$C_x(t) = \alpha_x C_x + (1 - \alpha_x) C_x(t-1) \quad (7)$$

$$C_y(t) = \alpha_y C_y + (1 - \alpha_y) C_y(t-1) \quad (8)$$

$$C_z(t) = \alpha_z C_z + (1 - \alpha_z) C_z(t-1) \quad (9)$$

$$\theta(t) = \alpha_\theta \theta + (1 - \alpha_\theta) \theta(t-1) \quad (10)$$

Hierbei sind die Werte der Gewichtungsfaktoren  $\alpha$  im allgemeinen zwischen 0 und 1. Auf diese Weise erhält man zeitlich geglättete Schätzwerte für die oben angegebenen Parameter. Im allgemeinen ist ein Wert von  $\alpha = 0,5$  zur Durchführung des Verfahrens gut geeignet.

Obwohl die angegebenen Berechnungen bereits meistens mit ausreichender Effizienz durchgeführt werden können, kann die Eignung des erfindungsgemäßen Verfahrens für die Echtzeitausführung noch durch weitere Verbesserungen merklich gesteigert werden. Diese möglichen Verbesserungen umfassen den Einsatz eines dynamisch gesteuerten Suchfensters und einer adaptiven Unterabtastung der Bilder.

Der Zweck eines dynamischen Suchfensters besteht darin, irrelevante Bereiche des Bildes von vornherein zu ignorieren. Dazu verwaltet das Verfahren vorzugsweise ein rechteckiges Suchfenster um die segmentierten Bildausschnitte. Für jedes nachfolgende Bild werden lediglich Bildbereiche innerhalb dieses Suchfensters für die Segmentierung bearbeitet. In jedem Iterationsschritt des geschilderten Segmentierungsalgorithmus werden die Ränder dieses Suchfensters für das folgende Bild wie folgt berechnet:

$$x_{\min} = \min_i (p_{ix}) - b_x$$

$$x_{\max} = \max_i (p_{ix}) + b_x$$

$$y_{\min} = \min_i (p_{iy}) - b_y \quad (11)$$

$$y_{\max} = \max_i (p_{iy}) + b_y$$

Auf diese Weise werden im folgenden Bild nur diejenigen Pixel durchsucht, die innerhalb des durch die Koordinaten  $x_{\min}$ ,  $y_{\min}$  und  $x_{\max}$ ,  $y_{\max}$  definierten Grenzen liegen. Die Konstante  $b$  (mit typischen Werten um 40) stellt sicher, daß das Fenster etwas größer ist als die tatsächlichen Ausmessungen der zu verfolgenden Hand.

Dabei ist es nicht erforderlich, jeden Bildpunkt innerhalb des Suchfensters zu untersuchen. In vielen Fällen ist dies auch für Echtzeitanwendungen zu zeitraubend. Um das Verfahren zu beschleunigen, wird in einem bevorzugten Ausführungsbeispiel des erfindungsgemäßen Verfahrens daher eine Unterabtastung des Bildes verwendet. Das bedeutet folgendes:

Wurde ein Bildausschnitt am Ort  $(x, y)$  untersucht, wird der nächste Bildausschnitt an der Stelle  $(x+s, y)$  ausgewählt. Wird dabei das Ende der Abtastzeile erreicht, wird statt dessen der Bildausschnitt, der bei  $(0, y+s)$  beginnt, untersucht. Dabei ist die Auswahl eines optimalen Unterabtastfaktors  $s$  wichtig aber nicht ganz einfach. Die Unterabtastung beeinflusst nämlich die Genauigkeit und auch die Geschwindigkeit der Verfolgung. Wählt man die Unterabtastung zu groß, dann werden die Lageschätzungen zu sehr verrauscht. Wählt man die Unterabtastung hingegen zu klein, wird das Verfahren insgesamt zu langsam und daher für eine Echtzeitverarbeitung ungeeignet.

Anstatt ein optimales Unterabtastverhältnis auf dem Wege des Probierens zu finden, verwendet das erfindungsgemäße Verfahren bevorzugt eine adaptive Technik zur automatischen Auswahl eines optimalen Unterabtastverhältnisses  $s$ . Dabei wird eine anzustrebende Bildfrequenz als Zielbildfrequenz vorgegeben. Der Segmentationsalgorithmus überwacht dabei laufend seine eigene Geschwindigkeit. Falls die Segmentationszeit mehr als 10% unter der gewünschten Segmentationszeit liegt, wird die Unterabtastung verringert, so daß eine größere Genauigkeit entsteht. Ist die Segmentationszeit hingegen 10% größer als gewünscht, dann wird die Unterabtastung erhöht um die Verarbeitungsgeschwindigkeit zu steigern. Dabei mißt das Verfahren jeweils nach 10 Bildern die Zeit, die für die Verarbeitung dieser 10 Bilder benötigt wurde. Ist diese Zeit  $t_p$  größer als das 1,1fache des Quotienten aus 10 und der Zielbildfrequenz, dann ist die Verarbeitungszeit zu klein und  $s$  wird erhöht. Andernfalls, wenn  $t_p$  kleiner als das 0,9fache des Quotienten aus 10 und der Zielbildfrequenz ist, ist die Verarbeitungszeit zu schnell und  $s$  wird erniedrigt.

Auf diese Art und Weise balanciert das Verfahren automatisch die sich widersprechenden Anforderungen zwischen Geschwindigkeit und Genauigkeit unter Berücksichtigung der Anforderungen des erfindungsgemäßen Verfahrens.

Ein wichtiger Gesichtspunkt dabei ist, daß bei einer Änderung des Unterabtastfaktors  $s$  die Zahl der Bildausschnitte in der Hand sich ändert. Aus diesem Grunde kann man nicht die tatsächliche Fläche eines Bildausschnitts für die Berechnung der Tiefeninformation verwenden. Z. B. wenn  $s$  verkleinert wird, gibt es mehr Bildausschnitte in der Hand, obwohl die Tiefe unverändert ist. Aus diesem Grunde wird der Faktor  $ap$  in Formel (6) gleich  $s^2$  gewählt. Hierdurch wird die Veränderung der Bildausschnittsfläche, d. h. der Fläche die eine Gruppe von